# Online Courses from Harvard and MIT

This dataset contains information about online courses, with columns including:

- **Institution**, **Course Number**, **Launch Date**, **Course Title**, **Instructors**, **Course Subject**, **Year**, **Participants**, and various statistics on engagement and demographics.

## 1. Identify the Entities and Create Tables

Based on the data, here are some logical tables you might consider:

- **Institutions**
- **Courses**
- **Instructors**
- **Course_Sessions** (details about each time a course is offered, such as year and engagement statistics)

## 2. Steps to Implement this Structure in MySQL

**Create the Tables**: Define each table in MySQL using CREATE TABLE statements, specifying primary and foreign keys to establish relationships.

```
CREATE TABLE Institutions (

    institution_id INT PRIMARY KEY AUTO_INCREMENT,

    institution_name VARCHAR(255) NOT NULL

);



CREATE TABLE Courses (

    course_id INT PRIMARY KEY AUTO_INCREMENT,

    course_number VARCHAR(255) NOT NULL,

    course_title VARCHAR(255),

    course_subject VARCHAR(255),

    institution_id INT,
```

```
    FOREIGN KEY (institution_id) REFERENCES Institutions(institution_id)
);

CREATE TABLE Instructors (
    instructor_id INT PRIMARY KEY AUTO_INCREMENT,
    instructor_name VARCHAR(255) NOT NULL
);

CREATE TABLE Course_Instructors (
    course_id INT,
    instructor_id INT,
    PRIMARY KEY (course_id, instructor_id),
    FOREIGN KEY (course_id) REFERENCES Courses(course_id),
    FOREIGN KEY (instructor_id) REFERENCES Instructors(instructor_id)
);

CREATE TABLE Course_Sessions (
    session_id INT PRIMARY KEY AUTO_INCREMENT,
    course_id INT,
    launch_date DATE,
    year INT,
    honor_code_certificates INT,
    participants INT,
```

```
    audited INT,
    certified INT,
    percent_certified DECIMAL(5, 2),
    percent_played_video DECIMAL(5, 2),
    percent_posted_in_forum DECIMAL(5, 2),
    percent_grade_higher_than_zero DECIMAL(5, 2),
    total_course_hours DECIMAL(10, 2),
    median_hours_for_certification DECIMAL(10, 2),
    median_age DECIMAL(4, 1),
    percent_male DECIMAL(5, 2),
    percent_female DECIMAL(5, 2),
    percent_bachelors_degree_or_higher DECIMAL(5, 2),
    FOREIGN KEY (course_id) REFERENCES Courses(course_id)
);
```

**3. Insert Data**:

- **Institutions, Courses, and Instructors**: Insert unique records into each main entity table.
- **Course_Sessions**: Insert session-specific data, linking each session to its respective course.
- **Course_Instructors**: Insert relationships between courses and instructors.

**1.** Institutions Table **:**

```
INSERT INTO Institutions (institution_id, institution_name)
VALUES
(1, 'MITx'),
```

```
(2, 'HarvardX'),
(3, 'StanfordX'),
(4, 'BerkeleyX'),
(5, 'YaleX'),
(6, 'ColumbiaX'),
(7, 'PrincetonX'),
(8, 'CornellX'),
(9, 'UPennX'),
(10, 'DukeX'),
(11, 'UChicagoX'),
(12, 'UMichiganX'),
(13, 'CaltechX'),
(14, 'BrownX'),
(15, 'DartmouthX');
```

2. **Courses Table**

```
INSERT INTO Courses (course_id, course_number, course_title, course_subject, institution_id)
VALUES
(1, '6.002x', 'Circuits and Electronics', 'Science, Technology, Engineering, and Mathematics', 1),
(2, '6.00x', 'Introduction to Computer Science and Programming', 'Computer Science', 1),
(3, '3.091x', 'Introduction to Solid State Chemistry', 'Science, Technology, Engineering, and Mathematics', 1),
(4, 'CS50x', 'Introduction to Computer Science', 'Computer Science', 2),
(5, 'PH207x', 'Health in Numbers: Quantitative Methods', 'Government, Health, and Social Science', 2),
(6, 'STAT101', 'Statistics for Data Science', 'Data Science', 3),
(7, 'BIO101', 'Introductory Biology', 'Biology', 4),
(8, 'ENG101', 'English Literature', 'Literature', 5),
```

```
(9, 'HIS101', 'World History', 'History', 6),

(10, 'ECON101', 'Principles of Economics', 'Economics', 7),

(11, 'PSY101', 'Introduction to Psychology', 'Psychology', 8),

(12, 'SOC101', 'Sociology Basics', 'Sociology', 9),

(13, 'PHIL101', 'Philosophy of Mind', 'Philosophy', 10),

(14, 'CHEM101', 'General Chemistry', 'Chemistry', 11),

(15, 'PHYS101', 'General Physics', 'Physics', 12);
```

3. **Instructors Table**

```
INSERT INTO Instructors (instructor_id, instructor_name)

VALUES

(1, 'Khurram Afridi'),

(2, 'Eric Grimson'),

(3, 'John Guttag'),

(4, 'Chris Terman'),

(5, 'Michael Cima'),

(6, 'David Malan'),

(7, 'Nate Hardison'),

(8, 'Rob Bowden'),

(9, 'Tommy MacWilliam'),

(10, 'Earl Francis Cook'),

(11, 'Marcello Pagano'),

(12, 'Andrew Ng'),

(13, 'Jennifer Widom'),

(14, 'Daphne Koller'),

(15, 'Michael Collins');
```

## 4.Course_Instructors Table

```
INSERT INTO Course_Instructors (course_id, instructor_id)
VALUES
(1, 1),
(2, 2),
(2, 3),
(3, 5),
(4, 6),
(4, 7),
(5, 10),
(6, 12),
(7, 14),
(8, 15),
(9, 13),
(10, 11),
(11, 9),
(12, 8),
(13, 4);
```

## 5. Course_Sessions Table

```
INSERT INTO Course_Sessions (
    session_id, course_id, launch_date, year, honor_code_certificates, participants,
    audited, certified, percent_certified, percent_played_video, percent_posted_in_forum,
    percent_grade_higher_than_zero, total_course_hours, median_hours_for_certification,
    median_age, percent_male, percent_female, percent_bachelors_degree_or_higher
)
```

**VALUES**

**(1, 1, '2012-09-05', 1, 1, 36105, 5431, 54, 54.98, 83.2, 8.17, 28.97, 418.94, 64.45, 26, 88.28, 11.72, 60.68),**

**(2, 2, '2012-09-26', 1, 1, 62709, 8949, 64, 64.05, 89.14, 14.38, 39.50, 884.04, 78.53, 28, 83.5, 16.5, 63.04),**

**(3, 3, '2012-10-09', 1, 1, 16663, 2855, 72, 72.85, 87.49, 14.42, 34.89, 227.55, 61.28, 27, 70.32, 29.68, 58.76),**

**(4, 4, '2012-10-15', 1, 1, 129400, 12888, 11, 11.11, 0.0, 0.0, 1.11, 220.9, 0.0, 28, 80.02, 19.98, 58.78),**

**(5, 5, '2012-10-15', 1, 1, 52521, 10729, 47, 47.12, 77.45, 15.98, 32.52, 804.41, 76.1, 32, 56.78, 43.22, 88.33),**

**(6, 6, '2013-01-05', 2, 2, 41200, 7351, 36, 50.21, 75.5, 11.0, 29.88, 315.24, 52.1, 24, 85.7, 14.3, 67.8),**

**(7, 7, '2013-02-14', 2, 3, 52300, 6542, 40, 45.6, 78.2, 12.1, 30.9, 392.7, 54.3, 27, 78.2, 21.8, 70.2),**

**(8, 8, '2013-03-18', 2, 1, 14300, 2011, 15, 40.5, 71.3, 10.5, 25.2, 287.43, 40.7, 29, 74.2, 25.8, 59.1),**

**(9, 9, '2013-04-02', 2, 1, 18921, 3212, 18, 43.5, 80.5, 13.2, 33.1, 317.94, 48.5, 31, 79.8, 20.2, 63.2),**

**(10, 10, '2013-04-20', 2, 1, 25420, 4188, 22, 42.3, 82.1, 15.3, 34.2, 354.56, 50.9, 28, 77.5, 22.5, 68.3),**

**(11, 11, '2013-05-07', 2, 1, 27530, 4321, 24, 46.2, 79.4, 12.8, 29.6, 374.24, 55.8, 26, 82.4, 17.6, 65.7),**

**(12, 12, '2013-05-25', 2, 2, 30042, 5276, 30, 48.9, 76.7, 11.9, 28.3, 290.2, 50.4, 27, 84.1, 15.9, 71.5),**

**(13, 13, '2013-06-15', 2, 1, 31253, 5943, 32, 49.1, 78.3, 13.5, 32.4, 420.7, 53.6, 29, 79.3, 20.7, 60.3),**

**(14, 14, '2013-07-03', 2, 1, 32940, 6131, 38, 53.4, 75.6, 12.9, 31.5, 408.7, 57.3, 24, 81.7, 18.3, 64.8),**

**(15, 15, '2013-08-12', 2, 1, 34718, 6220, 40, 55.2, 73.4, 14.0, 34.7, 435.8, 58.9, 25, 78.6, 21.4, 62.5);**